



Title	Study on Biologically Relevant Toxicogenomic Predictive Modeling using Machine Learning
Author(s)	永田, 圭佑
Citation	大阪大学, 2016, 博士論文
Version Type	VoR
URL	https://doi.org/10.18910/59587
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

論文内容の要旨

氏 名 (永 田 圭 佑)	
論文題名	Study on Biologically Relevant Toxicogenomic Predictive Modeling Using Machine Learning (機械学習による生物学的に妥当なトキシコゲノミクス予測モデルの研究)
論文内容の要旨	
本論文は、著者が大阪大学大学院工学研究科電気電子情報工学専攻に在学前及び在学中に行った機械学習による生物学的に妥当なトキシコゲノミクス予測モデルについての研究成果をまとめたものであり、以下の通り構成される。 第1章では、背景、課題、用いたデータソース、全章に共通する解析方法、及び論文構成について述べた。本論文の主目的は、正確かつ生物学的妥当なトキシコゲノミクス予測モデルを構築可能な機械学習法の探索である。この目的を達成するため、我々は「クラス関連ルールマイニング」「グループベースの構造正則化」「グラフベースの構造正則化」の3つの手法群について長所短所を明らかとすべく、マイクロアレイデータから肝重量増加を予測するトキシコゲノミクスモデルを構築し、それらの精度及び生物学的妥当性を従来法と比較した。 第2章では、クラス関連ルールマイニング法としてClassification Based on Association (CBA) 法を適用し、その精度と解釈容易性を線形判別法 (LDA) と比較した。CBAにより構築された識別器はLDAと比較し同等以上の精度を示し、解釈も容易であった。また、薬物代謝関連の遺伝子を中心に構成され、生物学的にも妥当であると考えられた。 第3章では、グループベースの構造正則化法としてLatent Group Lasso (LGL) 法を適用し、その精度と生物学的妥当性をLassoと比較した結果について述べた。このため、同一の転写制御因子下の遺伝子群を1つのグループと見なした。LGLはLassoと比較し同等の精度を有し、選択されたグループに基づいて毒性機序を考察できる点で優れていた。他方、LGLでは遺伝子レベルでスパース推定にならない点では課題が残った。 第4章では、グラフベースの構造正則化法としてGeneralized Fused Lasso (GFL) 法とGraph Lasso (GL) 法を適用し、その精度と生物学的妥当性をLasso及びLGLと比較した。このため、前章と同じグループ構造を、各転写制御と各下流遺伝子の間に辺を1つずつ設けることでグラフに再編成した。GFL及びGLともに精度はLasso及びLGLと同等であったが、GFLでは選択された遺伝子数がLGLよりも多くなったのに対し、GLではLasso並のスパースな推定が実現できた。また、GLでは選択された辺に基づいて毒性機序が考察可能な点でも優れていた。 第5章では、全章を通しての考察を加え、本研究で得られた成果を総括すると共に今後の展望を述べた。	

論文審査の結果の要旨及び担当者

氏 名 (永田 圭佑)			
論文審査担当者		(職)	氏 名
	主 査	教授	鷲尾 隆
	副 査	教授	馬場口 登
	副 査	准教授	河原 吉伸
	副 査	教授	滝根 哲哉
	副 査	教授	三瓶 政一
	副 査	教授	宮地 充子
	副 査	教授	井上 恭
	副 査	教授	駒谷 和範

論文審査の結果の要旨

本論文は、機械学習による生物学的に妥当なトキシコゲノミクス予測モデルについての研究成果をまとめたものであり、以下の通りの成果が示されている。

第1章では、背景、課題、用いたデータソース、全章に共通する解析方法、及び論文構成について述べている。本論文の主目的は、正確かつ生物学的妥当なトキシコゲノミクス予測モデルを構築可能な機械学習法の探索である。この目的を達成するため、「クラス関連ルールマイニング」、「グループベースの構造正則化」、「グラフベースの構造正則化」の3つの手法群について長所短所を明らかにすべく、マイクロアレイデータから肝重量増加を予測するトキシコゲノミクスモデルを構築し、それらの精度及び生物学的妥当性を従来法と比較している。

第2章では、クラス関連ルールマイニング法として Classification Based on Association (CBA) 法を適用し、その精度と解釈容易性を線形判別法 (LDA) と比較している。CBA により構築された識別器は LDA と比較し同等以上の精度を示し、解釈も容易であることを明らかにしている。また、薬物代謝関連の遺伝子を中心に構成され、生物学的にも妥当であることを示している。

第3章では、グループベースの構造正則化法として Latent Group Lasso (LGL) 法を適用し、その精度と生物学的妥当性を Lasso と比較した結果について述べている。同一の転写制御因子下の遺伝子群を1つのグループと見なしして解析を行った結果、LGL は Lasso と比較し同等の精度を有し、選択されたグループに基づいて毒性機序を考察できる点で優れていることを明らかにしている。他方、LGL では遺伝子レベルでスパース推定にならない点では課題が残る点を挙げている。

第4章では、グラフベースの構造正則化法として Generalized Fused Lasso (GFL) 法と Graph Lasso (GL) 法を適用し、その精度と生物学的妥当性を Lasso 及び LGL と比較している。前章と同じグループ構造を、各転写制御と各下流遺伝子の間に辺を1つずつ設けることでグラフに再編成して解析を行った結果、GFL 及び GL ともに精度は Lasso 及び LGL と同等であったが、GFL では選択された遺伝子数が LGL よりも多くなったのに対し、GL では Lasso 並のスパースな推定が実現できることを述べている。また、GL では選択された辺に基づいて毒性機序が考察可能な点でも優れることを明らかにしている。

第5章では、全章を通しての考察を加え、本研究で得られた成果を総括すると共に今後の展望を述べている。

以上のように本論文は、機械学習による生物学的に妥当なトキシコゲノミクス予測モデリングについて、予測精度と解釈容易性の両面から新しい方法論を提示している。そして、それによってもたらされた解析結果が、毒物が生体遺伝子ネットワークに影響を与える過程を、理解容易かつ高精度に明らかにすることを確認している。

よって本論文は博士論文として価値あるものと認める。